UCRL-JC-121201
PREPRINT
CONF-9508165--1

Contribution Towards Statistical Intercomparison of General Circulation Models

Jim Boyle*
and
Sailes Sengupta
PCMDI
Lawrence Livermore National Laboratory
Livermore, CA 94551

This paper was prepared for submittal to the
1995 Joint Statistical Meetings
Orlando, Florida
13–17 August 1995

June 1995

Lawrence Livermore National Laboratory

DISCLAIMER

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

CONTRIBUTION TOWARDS STATISTICAL INTERCOMPARISON OF GENERAL CIRCULATION MODELS

Sailes Sengupta and J. Boyle, PCMDI, Lawrence Livermore National Laboratory
Contact Author: Sailes Sengupta, L-264, LLNL, Livermore, CA 94551-9900

Keywords: general circulation model, model intercomparison, data reduction, principal components, neural networks, auto-associative neural networks, common principal components

1. Introduction and Background

The Atmospheric Model Intercomparison Project (AMIP) of the World Climate Research Programme's Working Group on Numerical Experimentation (WGNE) is an ambitious attempt to comprehensively intercompare atmospheric General Circulation Models (GCMs). The participants in AMIP simulate the global atmosphere for the decade 1979 to 1988 using a common solar constant and Carbon Dioxide(CO_2) concentration and a common monthly averaged sea surface temperature (SST) and sea ice data set. In this work we attempt to present a statistical framework to address the difficult task of model intercomparison and verification.

To begin we will attempt to summarize the aspects of the task of model intercomparison:

(a) We are required to compare a large number of models. (Some 30 modeling groups are participating in AMIP, for example.)

(b) The model output in each case is a multivariate vector of geophysical variables (temperature, wind, water vapor, etc.) with a large number of components. (The standard output of AMIP specifies some 15 variables.)

(c) Each component is defined over a spatial grid and hence is expected to have spatial autocorrelations of varying magnitudes. This issue is further complicated by the fact that the various models have different horizontal and vertical grids and thus may have different underlying correlation structures.

(d) The gridded data in each case has a temporal evolution based on the underlying physical processes and will in general have pronounced temporal autocorrelations.

An effective intercomparison/validation methodology must first devise a parsimonious representation of the spatio-temporal process(es) described above while providing a framework for intercomparison/validation.

A significant amount of effort has been expended by climatologists in the past to address this task. Bretherton et al.(1992) provided an intercomparison of several methods of analysis using the covariance structure. Preisendorfer and Mobley(1982 a-c).provided a comprehensive theory of data intercomparison by splitting the spatio-temporal 'distance' between two data sets into three parts based on the multidimensional geometric configuration of the data sets.Among other workers are Livezy (1985); Willmott et al.,(1985); Zwiers and Thibeaux (1987); Zwiers (1987); Wigley and Santer (1990).

Two other methods that have been attempted to understand this space-time structure are : multiple time series (MTS),Lutkepohl (1991) and a space-time stochastic process (STSP) approach with some representative works being Niu and Stein (1990), Kim and North(1992), Oshumi(1988), Martin (1990) and Basawa et al. (1990). In the search for an effective intercomparison tool, there are two conflicting requirements. Firstly, we need a parsimonious description of the data structure. Secondly, the results of the intercomparison should maintain the space-time structure of the original data, allowing for a physical interpretation of the results of the analysis. While the MTS approach meets the second requirement ,it is inadequate for the first.. A MTS based on a full global grid is difficult to analyze due to its computational complexity. Likewise, the STSP approach offers parsimony at the expense of detailed physical interpretation.In the pursuit of this dual objective, we may consider a reduced multiple time series(RMTS) where the number of components in a vector time series under consideration is significantly reduced. The most commonly used method for such a reduction is by means of a principal component analysis.

RMTS based on principal vectors: One may select a few significant principal vectors and study the temporal evolution of the corresponding principal components (PC) which are obtained as projection of the data field for each time instant onto the principal vectors.

Principal vectors use only linear functions of the variables. In many geophysical applications, the principal modes of a space-time process may not necessarily be expressible in terms of a linear set in the most parsimonious way. In such cases, one may use an auto-associa-

tive neural network (AANN) to extract nonlinear functions of the variables that may better summarize the data . In sec 2. we introduce briefly such networks and their use in data summarization..
While a RMTS derived by using either the PC or the AANN approach may be satisfactory for the understanding of a single model behaviour, it is inadequate for an intercomparison of several models and observations. In sec. 3 we motivate the use of common principal components (CPC) for such an intercomparison. Sec. 4 gives a description of the data and principal results. Subsection (a) describes the data and analysis showing an application of the nonlinear principal components by an auto-associative neural network and subsection (b) the same for an application of the method of common principal components. Finally, section 5 concludes with indications for future work.

2. Artificial Neural Network

Connection with Principal Component Analysis (PCA)

PCA is an essential technique in data compression, feature extraction, compact coding and computational efficiency.In the context of climatological data analysis the constraints and interdependency of spatio-temporal data can be identified and redundancy eliminated by the use of PCA. For example, the use of PCA is commonplace in climatological literature (Preisendorfer, 1988) and MPEG Video Compression is based on the same principle (LeGall, 1991).Briefly speaking, if x is a centered n-dimensional vector, PCA extracts p (<=n) linear combinations W x of the components of x where W is a p x n matrix of weights subject to the constraints that (a) the variance of each linear combination appearing as elements of W x is maximized and (b) the extracted linear combinations are mutually orthogonal.In practice one solves the eigensystem for the sample covariance matrix say, C with the resulting p dominant eigenvectors representing the principal vectors. Several workers in the Neural Network community have related (Bourlard and Kamp, 1987;Baldi and Hornik, 1989; Sirat, 1991, Oja, 1992) multilayer perceptron learning by back propagation algorithm with principal components extraction in classical statistics. Becker (1991) gives a comprehensive survey relating PCA models to unsupervised learning neural networks. While neural networks can derive the principal components, the self-organizing rules of PCA perform very poorly when outliers are present. Xu and Yuille (1995) discusses the problem and present robust self-organizing rules of PCA extraction based on statistical physics approach. In suggesting a transition from linear to 'nonlinear' principal components (sometimes called 'principal manifolds'), Demers and Cottrell (1992) argues that PCA finds an optimal linear subspace on which one projects the data with minimum loss of information (in the sense of 'explained' variance in the data). However, if the data lie on a nonlinear submanifold of the feature space, then the number of dominant PCs will overestimate the dimensionality. Take for example data from a sampled from a 3-d helix. The covariance will have a full rank (=3) with 3 distinct eigenvalue/eigenvector pairs leading to 3 distinct principal components. However, the intrinsic dimensionality of any 3-d curve such as a helix is one since it can ordinarily be parametrized by a single parameter.Furthermore, as has been observed by other authors (Kramer, 1991; Oja, 1991; Usui, Nakauchi and Nakano, 1991 and Demers and Cottrell, 1992), the addition of hidden layers between the inputs and the representation layer as well as between the representation and the output layer provide a network which is capable of learning nonlinear representation also.In the process, one achieves what may be termed a nonlinear analogue of PCA.In the following we present the sketch of such a network.(Fig. 1) The network consists of five layers which are fully interconnected. In addition to the input and output layers which are identical since the network is made to be auto-associative, we have a central representation layer (where the principal manifolds or nlpcs are generated as outputs) and two identical layers placed on the two sides of the representation layer.These last two layers as described above are called the encoding and decoding layer respectively. The training algorithm which is essentially adjustment of the connection weights to minimize the vector difference between the target and the network output is based on the back propagation of error algorithm.(Rumelhart and McClelland, 1986)

Network Architecture

Although there are no specific guidelines for the choice of the number of nodes in the encoding and the decoding layer, Kramer (1991) provides bounds based on the principle that the number of weights in the network should be a fraction of the number of constraints imposed by the data set. A few simplifying assumptions then lead to the constraints

$$M1 + M2 \ll n, M1 > f \text{ and } M2 > f$$

where M1, M2 are respectively the sizes of the encoding and the decoding layers, n is the size of the training set, and f is the number of nodes in the bottleneck layer.

In determining the size of the representation layer three different approaches have been indicated in the literature. Demers and Cottrell (1992) suggest a 'pruning' method based on successive elimination of representation nodes by penalizing variances. This results in encodings of minimum dimensionality with respect to allowable reconstruction error. Kramer (1991) on the other hand introduces a sequential determination of the nlpcs one at a time similar in spirit to its linear counterpart, namely the PCA extraction algorithm. Applied in this context, it amounts to using in a recursive manner the same network in fig. 2 except with a single node in the representation layer and in each recursion feeding the residual matrix obtained from the previous stage as the I/O pair. The residual matrix is simply the error matrix obtained by subtracting the output of the trained network from the input. The procedure stops when either a desired number of nlpcs have been extracted or a desired level of accuracy has been attained in the residual matrix.In yet a third method, instead of a sequential procedure, we may simply decide to use a fixed number p say, of nodes in the representation layer and extract the p nlpcs. This will however preclude any ranking of these nlpcs. In our work, we have used the last two methods only.

3. Common principal components (CPC): A tool for studying common covariance structure

To motivate the idea of common principal components in the context of climate models/observation intercomparison,we briefly indicate here how PCs are normally used in the intercomparison of two or more temporally evolving climatological fields. We first compute the dominant principal vectors (ranked by eigenvalues) for the data fields under comparison. Then intercompare the time series resulting from the projection of the data on the principal vectors. We also intercompare the fields of ranked principal vectors of each data set with their counterpart in another data set. A potentially serious problem in this approach is that the principal vectors for different fields under comparison, ordered by the respective principal values, may not necessarily be representing the same physical processes (when such underlying processes can be identified)in that order. It is this difficulty that motivated us to seek a common frame of reference, the *common principal vectors,* for the purpose of intercomparison of such fields. The common principal vectors identify the principal (spatial)modes of variation that are common to the data sets under comparison. For a detailed model intercomparison, one needs also to study the similarities in the *temporal evolution of the data sets along the identified common principal modes.* To do this ,we will look at the projection of all data sets on each of the common principal vectors getting as many sets of time series as the number of dominant common principal vectors. Each such set consists of one time series for each model under intercomparison as well as one for the observations.

The discussion above leads to the following line of enquiry. First, given two or more multi-dimensional data sets does their exist a common set of orthogonal eigenvectors? Second, if a complete set of common eigenvectors does not exist, is there a partial one? The questions can be addressed in a more general context by considering the covariance structures of these data sets. Specifically, let there be k fields under comparison with p components each. Let S_i, i=1,2,. k, be their respective sample covariances. One of the most important questions in the intercomparison of these fields is: Are the corresponding population covariance matrices $\sum_i$ similar in any meaningful way? Flury(1988) has provided the following levels of hierarchy of similarity of covariance matrices: $\sum_i$, i=1,2, ... k. : equality, proportionality, having common eigenvectors and having partially common set of eigenvectors. In this paper, we shall restrict ourselves to the third level, namely the commonality of the eigenvectors.One major advantage in using the CPC model is that one can compare *corresponding* principal components. A formal test of significance for the hypotheses of (partial) commonality of the principal axes of representation of two(or several) fields of data along the line given in Flury(1988) is however not possible to implement directly. These tests of significance require that the sample fields (over discrete time instants) be independent. This, in most cases, is not a valid assumption since fields over successive time instants are in fact correlated. This problem itself does not preclude the use of common principal components as a diagnostic tool for understanding the commonality of the fields.

Time series prediction and intercomparison

One way of comparing two (or more) time series resulting as the common principal components is to check how well the identification of parameters in one can be used to predict the other. This notion of predictability of one series in terms of the other can be extended one more step by regarding the two series in a bivariate context. More precisely, one may consider one series as a (linear)filtered version of another and estimate in an 'optimal' way the filter coefficients(Newton 1988). One may expect to do a little better if one allows for nonlinearity in the filter. We indicate below an implementation of this process in a a nonparametric setting by using an ANN directly for the prediction, bypassing the need to identify the model first as in the ARMA approach and avoiding the linearity restriction in the latter.

The fundamental problem in a one-step prediction of one time series in terms of another can simply be stated as the estimation of a mapping f as in

$$Y(t+n)=f(X(t),X(t+1),...,X(t+n)) \quad (1)$$

where X(T), Y(T), T = t, t+1, ..., t+n denote respectively the values of two time series at time T. In the context of the problem of model intercomparison, one can look at two time series X(T), Y(T) of a specified CPC pair resulting from two model outputs(or a model output and observations) and find a predictive function of the form (1) for the Y(T) series. This function is encapsulated in the form of weights of the ANN trained by 'examples' selected as time segments of fixed length n from the series X(T). These weights are analogous to the regression coefficients in a regression model. The input (X(t), X(t+1), ... , X(t+n-1)) is a segment of length n, and the output is Y(t+n-1) for t=1,2,..,N-n+1. The inputs are taken from the time series segment X(1),X(2),...,X(N) . This estimated function can then be used to predict Y(t+n-1) based on an input segment (X(t), X(t+1),..., X(t+n-1)) for different values of t. The corresponding ANN output is the predicted value of the Y(t) series based on the X(t) series. A widely used measure of skill of a predictor is the correlation coefficient R between actual and predicted values (Anthes 1984). This or other measures of predictive skill can then be used to validate the similarity of the two models(or the model and the observations). The process can of course be repeated for the comparison of all leading CPCs.

An overall intercomparison strategy

Common principal vectors (CPV) provide a means of representation of data from multiple fields in a common frame of reference. They in turn lead to groups of time series of principal components where the time series within each group need to be intercompared. A method of intercomparison by traditional methods as well as methods based on ANN has been introduced in the last paragraph. Now we combine the two steps to outline an overall strategy of space-time field intercomparison.

1. For a given meteorological/oceanographic variable over k space-time fields, determine if the mean fields under comparison are very much alike. If they are, one may consider the comparison of anomaly fields. For example, one might consider monthly temperature anomaly patterns over a certain period as given by observation, GCM output, or analysis derived from an incomplete set of observations.
2. Compute the common principal vectors based on the covariance matrices under study retaining only the dominant ones based on some heuristic criterion such as the size of the eigenvalues.
3. For each CPV, project each of the k space-time fields on the CPV by taking the scalar product of the CPV and the sample field at each time instant to get k univariate time series.
4. Using the methods indicated earlier in this section, the k univariate time series can now be intercompared with regard to their evolutionary patterns.

Thus, in the search for a coupled set of patterns one first looks at the similarity in the spatial distribution through the orthogonal common principal vectors and then in the temporal characteristics of the derived principal components. The latter task can be accomplished by traditional methods based on linear regression or nonlinear methods based on artificial neural networks.

In the context of *model intercomparison*, two data sets (model/model or model/observation) would be considered as 'similar' with increasing degree of similarity in the order indicated below, if:

(i) The significant common principal components within each pair under comparison explain a 'large' portion of the variations in the fields under comparison.

(ii) A high degree of predictive skill is demonstrated when one of the series is used in the prediction of another. A widely used measure of skill of a predictor is the correlation coefficient R between actual and predicted values (Anthes 1984)).

4. Data and Results

(a) Data for neural network

The data used in this part of the study consists of precipitation observations gridded to a 4degree by 5 degree latitude, longitude lattice. The observations are from surface stations over land and satellite MSU estimates, Spencer(1993), over the oceans. The bulk of the analysis grid used here is over the United States where the observational network provides reliable precipitation fields. The data are monthly averages for the 120 months from January 1979 to December 1988. The 120 month mean was subtracted from each gridpoint to form the deviations. The seasonal cycle was retained since it was of interest to compare how well the GCMs simulated this cycle. The data comprised a matrix of 120 time points at 95 space points. Figure 3 shows the spatial coverage of the data.

Analysis:

The PC analysis used the standard routine, PRINC from IMSL (1994) to compute the principal components from the covariance matrix computed from the data consisting of the 120 time points at 95 space points.

The same data were used as input to the neural net described above.

Results:

The results will focus on the first components of each method since these represent the seasonal cycle and allow a fairly direct interpretation in the limited space available in this paper.

Figure 3 shows the projection of the first nlpc onto the data. This spatial distribution allows for some physical insight into the components. The projection for the first PC was very similar overall, with some differences in detail. From the times series in Figs. 4 and 5. it can be seen that these first components are a representation of the seasonal cycle. Figure 3 clearly shows the characteristic west and east coast wintertime maxima in precipitation due to cyclonic storms, while the mid-continent has a summertime maximum which is attributable to convective, severe storms.

Both the PC and nlpc techniques capture the bulk of the seasonal oscillation but the nlpc displays a somewhat sharper distinction and transition from the winter to summer precipitation regimes. This is shown in the spectra of Figs. 6 and 7. The PC analysis has a significant contribution from the second PC at 12 months (1/ 12 = 0.083) while the nlpc has virtually all the seasonal contribution in the leading component. The extra freedom allowed by the non-linearity permits this more distinct characterization. This property would be of use when the comparisons are made with GCM output. The proper simulation of the seasonal cycle is an essential benchmark for an acceptable integration.

On the other hand, the close resemblance of the analogous PC figure(not shown)to Fig. 3 indicates that the linear PC does capture the essence of the spatial distribution and might well be adequate for most purposes.

(b Data description - CPC

The data sets used for the example application of the CPC methodology are the 200 hPa velocity potential fields from two sources. The 200 hPa velocity potential is a scalar potential for the divergent component of the horizontal wind. The divergent wind is directly linked to the upward motion in the atmosphere and thus to regions of precipitation. The 200 hPa level is located at about 12 km in the atmosphere. This is approximately the level of strong outflow from deep tropical convective storms. One set of data are from the monthly mean wind fields from five simulation of ten years in duration. Each of these simulations was for the decade 1979 to 1988. The model used was the ECMWF GCM cycle 36. This is an atmospheric GCM so the sea surface temperatures must be specified. The model has 19 levels in the vertical and a horizontal resolution of T42. The model is in almost all respects the same as that described by Miller et al. (1992). The sea surface temperatures are specified in accordance with the AMIP guidelines (Gates, 1992). The surface land temperatures are allowed to vary in accordance to the surface parameterizations employed.

In these five cases the observed sea surface temperatures for the decade were used. The integrations only varied by the initial conditions used, all boundary forcings and other external parameters were identical. The initial conditions for the first run were the observed data for 1 Jan 1979. The initial conditions for the subsequent runs were taken from the ending state of the previous run.In this decade The atmosphere has been shown to be chaotic in the sense that the simulations are sensitive to the initial conditions. However, in the experiment described here the specification places a strong constraint on the path of the simulation in phase space. It is known that the specified SST will leave an imprint on the simulations, what needs to be determined is to what extent the various simulations have a common component, presumably due to the boundary forcing by the SSTs. There will also be a component due to the seasonal changes of insolation but in this analysis the seasonal cycle has been removed from the data.

The decade 1979 to 1988 had two very prominent El Nino / Southern Oscillation (ENSO)events in 1982/ 83 and 1986/87. These events are manifested by an extrusion from the South America coast of anomalously warm SSTs in the tropical Pacific. It is well documented that these warm ocean temperatures lead to an enhancement of tropical convection and precipitation in the Tropical Eastern Pacific. The 200 hPa divergent wind will be a useful measure of the dynamic response of the model atmosphere to the varying SST.

The monthly mean velocity potential fields for the 5 simulations had their seasonal cycle removed. The model output the data on a global grid of 128 longitudinal points and 64 latitudinal points. These data were transformed to the orthogonal spherical harmonics nad the spherical harmonic series was truncated at T10. This limits the results to large scale features but allows a fit in the spatial domain since there are 110 spatial coordinates (110 coefficients of the spherical harmonics decomposition) and 120 time points. From these data were then formed the covariance matrices for input into the CPC algorithm.

The algorithm used for determining the common eigenstructure was that of Flury and Gautschi (1986). The code was tested against the IMSL (1991) routine KPRIN and the results were identical. The IMSL routines were not used because we wanted to have access to some intermediate results and the IMSL routines were unable to permit this.

Table 1 : Percent variance explained by CPCs for the 5 simulations.

run	CPC 1	CPC 2	CPC 3
1	37	31	7
2	39	28	9
3	41	25	9
4	38	28	7
5	43	24	8

Results - CPC

Figure 8 is a plot of the Southern Oscillation Index (SOI) from the Climate Analysis Center. This index is the difference in atmospheric pressure between Darwin, Australia and Tahiti. It is tightly linked to the cycle of SST and the atmospherics response the SST. The two distinct dips in 1982/83 and 1986/87 represent two strong ENSO events, the 82/83 event being the strongest on record. The smooth curve in Fig. 8 is the result of an 8 point gaussian filter. This filter is often used by the CAC to emphasize the component forced by the SST variations which have somewhat longer timescales than the atmospheric variations. Figure 9 present the leading CPC for the 200 hPa velocity potential five simulations. Table 1 gives the precent variance explained by the leading 3 CPCs. All the simulations share the same leading three CPCs. Two things are evident. First, the simulations are all follow a similar time evolution which clearly reflect the pattern of the ENSO activity for the decade. This shows the influence of the SSTs common for all the simulations. Second, the curves are not identical. There is quite a bit of variation especially in the first three years. This is an indication of the chaotic noise. Notice, that after the onset of the first ENSO event in 1982 there is a better phase locking of the simulations. Figure 9 is a geographical plot of the divergence pattern of the leading CPC. It should be noted that most of the amplitude of the signal is over the Tropical Pacific and the pattern is broadly consitent with that expected from observed precipitaion anomalies associated with ENSO events.

Figure 10 used the same data as in Fig. 8 except that is shows the differences in each simulation from the mean of all the simulations at each time point. These curves are in a sense a measure of the non-deterministic component of the flow. The data are quite variable but the mean amplitude remains relatively constant over the decade with little obvious dependence on the ENSO cycle. There is also some hint at a systematic component of the differences linked to the magnitude of the differences in the initial conditions. The models maintain the relative ordering of the first month (which would he most strongly influenced by the initial conditions) for the entire 120 month length of the simulations.

The points to be made here is not the discovery of new relatinships but a measure of the impact of the SSTs on the simulations with varying initial conditions. Such, ensemble integrations are now commonplace among the major weather forecasting centers. The CPC methodology provides a framework for combining these ensembles into a single field. This combining is necessary since the number of members of the ensemble is often greaten than 20. This provides more information than can be easily used by a forecaster.

There is some indications that the SST are predictable a month or a season in advance. The large time scales associated with the ocean compared to the atmosphere makes this possible. If the atmospheric models are then driven by these SSTs, then a climate prediction can be made. A CPC analysis of an ensemble of such atmospheric predictions would be an efficient way of producing a robust climate forecast.

The figures indicate that beyond the ENSO signal the model and the observations atmosphere do not have a great deal in common. Each has a different response given a common SST forcings of the decade. This is not unexpected since on the global scale a great many more variables influence the variability of the integrated temperature field besides the SSTs. The CPC analyses does allow this difference in the model and observation to be seen clearly. The CPC approach allows one to see that the ENSO response plays a greater role in the model than the atmosphere. In addition the approach clearly shows the two data sets have little in common beyond the ENSO response. One does not have to hunt through fields of PCs looking for similar or dissimilar components, the CPC technique has essentially done this in a convincing fashion. Figures 10b,c,d graphically show that the two data sets evolve through the decade with little correspondence.

5.Conclusions and further research

A parsimonious representation of the spatio-temporal data derived from the observations or as model outputs is a necessary first step in understanding the complex data sets obtained as model output or from observations. To achieve this, in addition to the traditional method based on principal components, we have considered the use of AANN as a possible tool. For model validation/intercomparison on the other hand a straightforward application of the PCA does not seem to work as well as the CPC based approach especially when there are several models under comparison.The latter starts with a reduced data set in the form of a MTS consisting of a few orthogonal *common* principal components for the data sets under comparison. The use of CPCs are not limited to model intercomparisons only. In fact, it is a powerful tool in detecting coupled patterns involving several simultaneous spatio-temporal fields of meteorological variables. This last feature makes it a potentially valuable tool in understanding the physical processes associated with these fields. We have indicated briefly a method based on ANN that is capable of being used for the intercomparison of the PCs resulting from different models as time series. Developing a test similar to Flury (1988) when the data fields are temporally correlated will help determining the statistical significance of the commonality

Acknowledgments. The cooperation of the ECMWF in making their forecast model available and in providing expert technical advice for this research is gratefully acknowledged. This work was performed under the auspices of the Department of Energy Environmental Sciences Division by the Lawrence Livermore National Laboratory under contract W-7405-ENG-48

References

Anthes, R.A., 1984: *Predictability of Mesoscale Meteorological Phenomena, Predictability of Fluid Motions*, G. Holloway and B.J. West (Eds.), Amer. Inst. Phys., 247-270.

Baldi Pierre and Kurt Hornik, 'Neural Networks and Principal Component Analysis: Learning ffrom Examples Without Local Minima,' Neural Networks, Vol 2, pp 53-58, 1989

Basawa, I.V., P.J. Brockwell and V. Mandrekar, 1990: Statistical Analysis of Separable ARMA Processes on a Lattice. Tech. Reprt., The University of Georgia.

Becker, S., 'Unsupervised Learning Procedures for Neural Networks,' International JOurnal of Neural Systems, 2, 17-33, 1991.

Bourlard, H and Y. Kamp, 'Auto-association by Multi-Layer Perceptron and Singular Value Decomposition,' Biological Cybernetics, 59, 291-294, 1988.

Bretherton, C.S., C. Smith and J. Wallace, 1992: An intercomparison of methods for finding coupled patterns in climate data. *J. Clim.*, **5**, 541-560.

Daley, R. (1991) *Atmospheric Data Analysis*, Cambridge University Press, Cambridge, 457pp.

DeMers, David and Garrison Cottrell, ' Non-Linear Dimensionality Reduction,' in Advances in Neural Processing Systems 5, eds Hansen et al.Morgan Kaufmann, 1992.

Flury, B. ,and Gautschi, W., 1986: An algorithm for simultaneous orthogonal transformation of several positive definite symmetric matrices to nearly diagonal form. *Siam.Jour. Sci. and Statist. Comp.*, **7**, 169-184.

Flury, B., 1988: *Common Principal Components and Related Multivariate Models*, J. Wiley, New York, pp 258.

Gates, W. L., 1992: AMIP: The atmospheric model intercomparison project. *Bull. Amer. Meteor. Soc.*, 73, 1962-1970.

HIMSL Stat/Library, 1991: IMSL Inc., Houston, 1578pp.

Johansson, E. M., F. U. Dowla, and D. M. Goodman, 1992: Backpropagation learning for multilayered feedforward neural networks using the conjugate gradient metho. *International Journal of Neural Systems* **2**, 291-301.

Kramer, Mark A., 'Nonlinear Principal Comp[onent Analysis Using Autoassociative Neural Networks,' AIChE Journal, vol 37, No. 2, 1991.Lutkepohl, H., 1991: *Introduction to Multiple Time Series Analysis*. Springer Verlag, New York, 545pp.

LeGall, D., MPEG: A Video Compression Standard for Multimedia Applications,' Communications of the ACM, 34, 46-58, 1991.Martin, R.J., 1990: The use of time series models and methods in the analysis of agricultural field trials. *Commu. Statist. Theory Meth.*, **19(1)**, 55-81.

Miller, M. J., A. C. M. Beljaars and T. N. Palmer, 1992: The sensitivity of the ECMWF model to the parameterization of evaporation from the Tropical oceans. *J. Climate*, **5**, 418-434.

Niu, X., and M. Stein, 1990: Space-Time ARMA Models for Satellite Ozone Data, Interface '90, *Proceed-*

ings of the 22nd. Symposium on the Interface, May 16-19, 225-234.

LeGall, D., MPEG: A Video Compression Standard for Multimedia Applications,' Communications of the ACM, 34, 46-58, 1991

Oja, Erkki, 'Principal Components, Minor Components, and Linear Neural Networks,' Neural Networks, Vol 5. pp 927-935, 1992

Oshumi, A., 1988: Linear estimation of random fields with second order increments and its application. *Automatica*, **24**, 203-215.

Preisendorfer, R.W., and C.D. Mobley, 1982 a: *Data Intercomparison Theory I*, NOAA Tech. Memo ERL PMEL-38, 48 pp.

Preisendorfer, R.W., and C.D. Mobley, 1982 b: *Data Intercomparison Theory II*, NOAA Tech. Memo ERL-PMEL-39, 94pp.

Preisendorfer, R.W., and C.D. Mobley, 1982 c: *Data Intercomparison Theory III*, NOAA Tech. Memo ERL-PMEL-39, 123 pp

Preisendorfer, R W, 'Principal Component Analysis in Meteorology and Oceanography,' Developments in Atmospheric Science 17, 425pp., Elsevier, Amsterdam, 1988.Weare, B. C., A. R. Navato and R. E. Newell, 1976: Empirical orthogonal analysis of Pacific Sea surface temperature. *J. Phys. Oceanogr.*, **6**, 671-678.

Rumelhart, D. and James McClelland, Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol 1, MIT Pr., Cambridge, MA 547 pp., 1986

Weare, B. C., A. R. Navato and R. E. Newell, 1981: Autoregressive modelling and spectral analysis of array data in the plane. *IEEE Trans. Rem. Sens.*, **GE-19**, 15-23.

Weare, B. C., A. R. Navato and R. E. Newell, 1983: Statistical spatial series modelling II: some further results on unilateral lattice processes. *Adv. Appl. Prob.*, **15**, 562-584.

Wigley, T.M., and B.D. Santer, 1990: Statistical comparison of spatial fields in model validation, perturbation and predictability experiments. *J. Geophys. Res.*, **95** 851-865.

Willmott, C.J., S. G. Ackleson, R. E. Davis, J. J. Feddema, K. M. Klink, D. R. Legates, J. O'Donnel and C. M. Rowe, 1985: Statistics for the evaluation and comparison of models. *J. Geophys. Res.*, **90**, 8995-9005.

Xu, Lei and Alan Yuille, 'Robust Principal Component Analysis by Self-Organizing Rules Based on Statistical Physics Approach,' IEEE Trans. Neural Networks, Vol.6, No. 1, 131-143, 1995Zwiers, F.W., 1987: Statistical considerations for climate experiments, Part II: multivariate tests. *J.Clim.Appl. Meteor.*, **26**, 477-487.

Zwiers, F.W., and H.J. Thiebaux, 1987: Statistical considerations for climate experiments, Part I: scalar tests. *J.Clim.Appl. Meteor.*, **26**, 464-476.

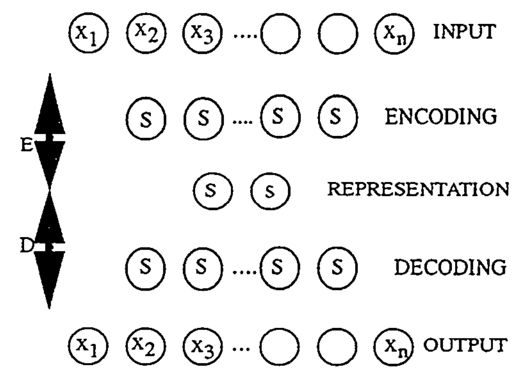

Figure 1 A fully interconnected (connections not shown) auto-associative neural network. The representation layer is shown with two (nlpc) nodes.

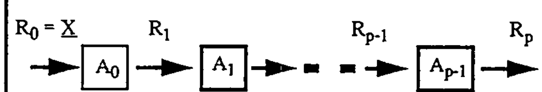

Figure 2. Sequential selection of non-linear principal components showing a sequence of p auto-associative networks A_0, A_1, ...A_{p-1} with each A_i configured like A in Fig. 1 (except with a single node in each representation layer). The I/O pair within each A_i during training are the successive residuals R_0, R_1, ...R_{p-1} with the original data matrix in R_0.

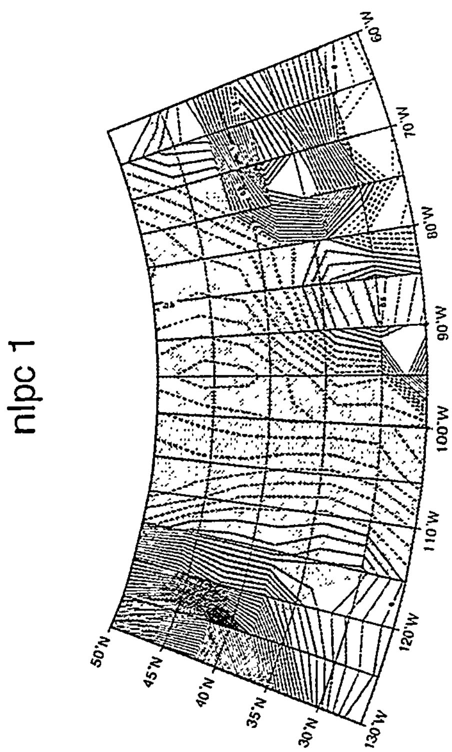

Figure 3. Projection of the first nlpc. The data is the observed precipitation. The analysis grid is a 4 x 5 latitude longitude grid over the US. Dashed lines indicate negative deviations, solid lines positive deviations.

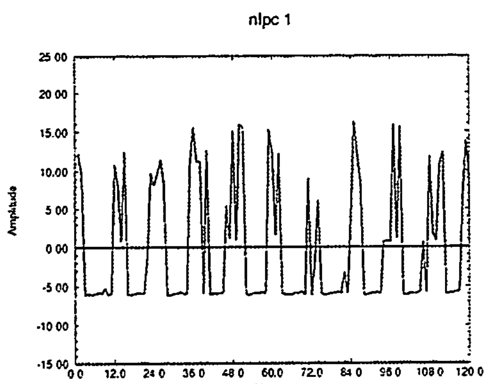

Figure 4. First nlpc of precipitation data.

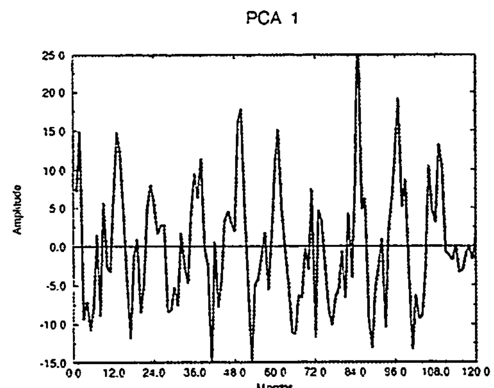

Figure 5. First PC of precipitation data.

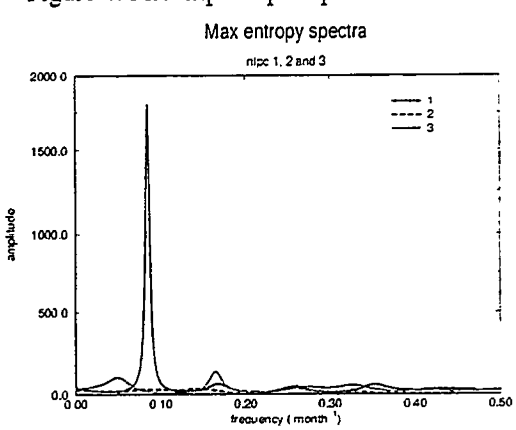

Figure 6. Spectra of first nlpc using maximum entropy estimation.

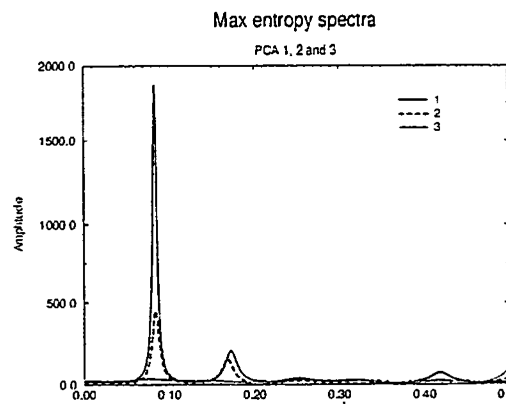

Figure 7. spectra of first PC using maximum entropy estimation

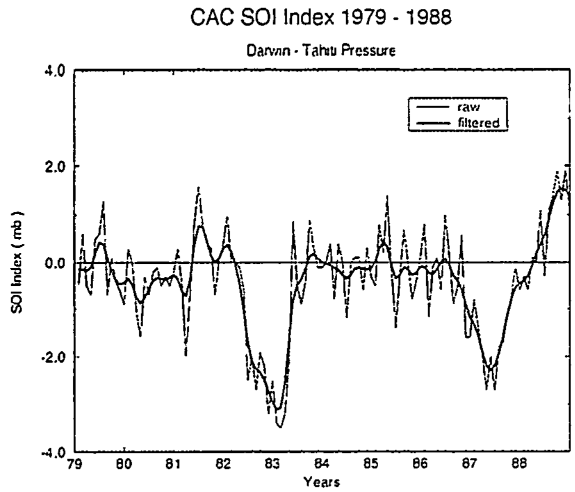

Figure 9 . Time series of the Southern Oscillation Index.

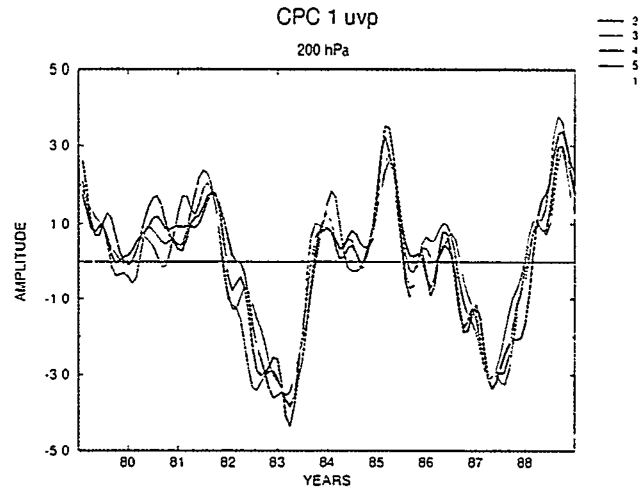

Figure 10. Leading CPC for the five ensmble simulations for the 200 hPa velocity potential.

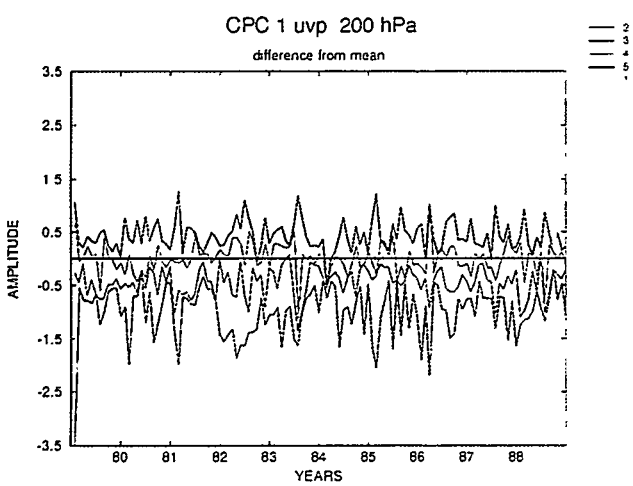

Figure 11 The difference from the mean for the five curves shown in Fig. 2.

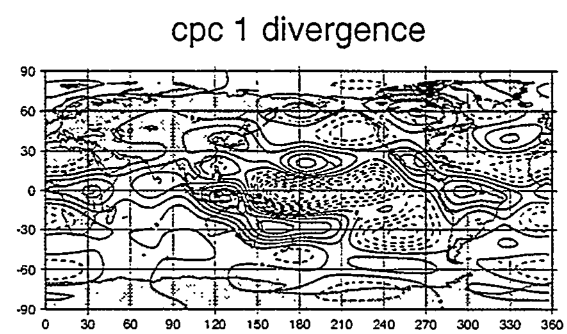

Figure 12. The leading CPC for all five simulations. The dashed contours indicate anomalous divergence for times when the curve in Fig. 1 is negative.